Raymond Lin
304937942

CS M146 - Problem Set 0

1. $y = z \sin(x) e^{-x}$

$$\boxed{\frac{\partial y}{\partial x} = z \cos(x) e^{-x} + -z \sin(x) e^{-x}}$$

2. $X = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$ $y = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ $z = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

a) $y^T z = ?$

$y^T = [1 \quad 3]$

$y^T z = [1 \quad 3] \begin{bmatrix} 2 \\ 3 \end{bmatrix} = (1 \times 2) + (3 \times 3) = 2 + 9 = \boxed{11}$

b) $Xy = ?$

$$Xy = \underset{2\times 2}{\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} 1 \\ 3 \end{bmatrix}} = \underset{2\times 1}{\begin{bmatrix} (2\times 1)+(4\times 3) \\ (1\times 1)+(2\times 3) \end{bmatrix}} = \boxed{\begin{bmatrix} 14 \\ 7 \end{bmatrix}}$$

c) By Gauss Jordan:

$X = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \leftarrow$ No

By determinant:

$\det(X) = ad - bc = (2\times 2) - (1 \times 4) = 0 \leftarrow$ No

$\boxed{X \text{ is not invertible}}$

d) $RREF(X) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ $\boxed{\text{rank}(X) = 1}$

3. a) $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(1+1+0+1+0)$

$= \frac{3}{5}$ $\boxed{\mu = \frac{3}{5}}$

b) $s^2 = \frac{\Sigma(x_i - \mu)^2}{n-1}$

$= \frac{(1-\frac{3}{5})^2 + (1-\frac{3}{5})^2 + (-\frac{3}{5})^2 + (1-\frac{3}{5})^2 + (-\frac{3}{5})^2}{5-1}$

$= \frac{2(\frac{9}{25}) + 3(\frac{4}{25})}{4}$

$= \frac{18+12}{25} \cdot \frac{1}{4} = \frac{\cancel{30}^{\cancel{6}\,3}}{\cancel{25}_5} \times \frac{1}{\cancel{4}_2} = \frac{3}{10}$ $\boxed{s^2 = \frac{3}{10}}$

* c) Probability of 3 Tails

$P(X=3) = \binom{n}{k} p^k (1-p)^{n-k}$

$= \binom{5}{3} 0.5^3 \cdot 0.5^2 = \frac{5!}{3!\,2!} \cdot 0.5^5$

$= 0.3125$

Probability of 3 Tails in that order $(1,1,0,1,0)$

- total possible outcomes $= 2^5 = 32$

- $P(1,1,0,1,0) = \frac{1}{32} = 0.03125$

d) $\ln\left(\binom{n}{k} p^k (1-p)^{n-k}\right)$

$= \ln\binom{n}{k} + \ln(p^k) + \ln((1-p)^{n-k})$

$= \ln\binom{n}{k} + k\ln(p) + (n-k)(\ln(1-p))$

$\frac{d}{dp}\left(\ln\binom{n}{k} + k\ln(p) + (n-k)(\ln(1-p))\right) = 0$

$= \frac{k}{p} + \frac{1}{1-p} \cdot (n-k) \cdot -1 = \frac{k}{p} + \frac{k-n}{1-p}$

$$\frac{k}{p} + \frac{k-n}{1-p} = 0 \qquad \frac{k}{p} = \frac{n-k}{1-p}$$

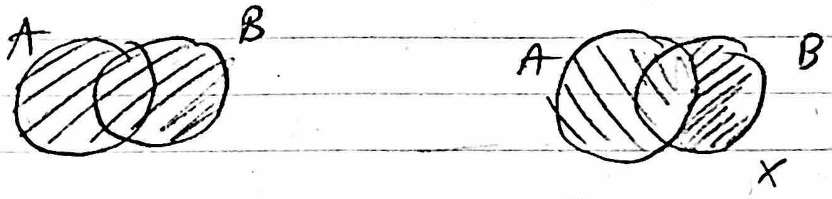
$$k(1-p) = p(n-k)$$

$$k - \cancel{kp} = np - \cancel{kp}$$

$$\boxed{p = \frac{k}{n} = \frac{3}{5}} \; \square$$

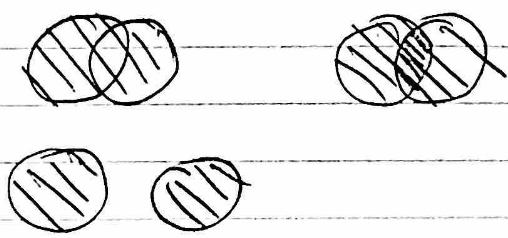
e) $p(X=7 \mid Y=6) = \dfrac{0.1}{0.25} = \boxed{0.4}$

4. a) $P(A \cup B) = P(A \cap (B \cap A^c))$



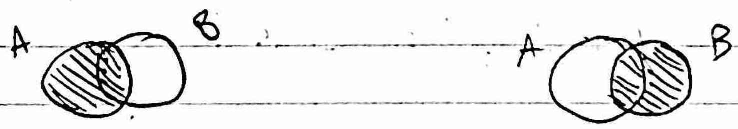
$\boxed{\text{False}}$

b) $P(A \cup B) = P(A) + P(B)$



False, unless A & B are mutually exclusive

c) $P(A) = P(A \cap B) + P(A^c \cap B)$



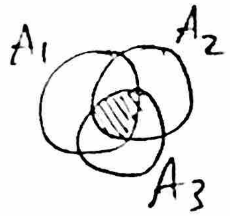
$\boxed{\text{False}}$

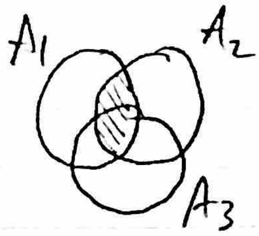
d) $P(A \mid B) = P(B \mid A)$

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

False, unless $P(A) = P(B)$

e) $P(A_1 \cap A_2 \cap A_3) = P(A_3 \mid (A_2 \cap A_1)) \times P(A_2 \mid A_1) \times P(A_1)$





$$P(A \cap B) = P(A \mid B)\ P(B)$$

$$\begin{aligned}
P(A_1 \cap A_2 \cap A_3) &= P((A_1 \cap A_2) \cap A_3) \\
&= P((A_1 \cap A_2) \mid A_3) \times P(A_3) \\
&= P(A_3 \mid (A_1 \cap A_2)) \times P(A_1 \cap A_2) \\
&= P(A_3 \mid (A_2 \cap A_1)) \times P(A_1 \mid A_2) \times P(A_2) \\
&= P(A_3 \mid (A_2 \cap A_1)) \times P(A_2 \mid A_1) \times P(A_1)
\end{aligned}$$

$\square$

$\boxed{\text{True}}$

5. a) Gaussian $\Rightarrow$ ii) $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x-\mu^2}{2\sigma^2}}$

b) Exponential $\Rightarrow$ iv) $\lambda e^{-\lambda x}$ when $x \geq 0$, 0 otherwise

c) Uniform $\Rightarrow$ v) $\frac{1}{b-a}$ when $a \leq x \leq b$, $\emptyset$ otherwise

d) Bernoulli $\Rightarrow$ i) $p^x(1-p)^{1-x}$, when $x \in \{0,1\}$, $\emptyset$ otherwise

e) Binomial $\Rightarrow$ iii) $\binom{n}{x} p^x (1-p)^{n-x}$,

6. a) $\mu = E(\underline{X}) = \sum_x x p(x)$ $\qquad p(x) = p^x(1-p)^{1-x}$

$= \cancel{0 \cdot p^0 (1-p)^{1-0}}^{\,0} + 1 \cdot p^1 (1-p)^{1-1}$ $\qquad x \in \{0,1\}$

$= p(1-p)^0$

$= p \cdot 1$

$= p$ $\qquad$ Thus $\mu = p$

$Var(\underline{X}) = E(\underline{X}^2) - (E(\underline{X}))^2$

$$Var(\underline{X}) = E((\underline{X}-\mu)^2)$$

$$= \sum_x (x-\mu)^2 p(x)$$

$$= (0-\mu)^2 p^0 (1-p)^{1-0} + (1-\mu)^2 p^1 (1-p)^{1-1}$$

$$= \mu^2 (1-p) + p(1-\mu)^2$$

$$= \mu^2 - \mu^2 p + p(1-2\mu+\mu^2)$$

$$= \mu^2 - \mu^2 p + p - 2\mu p + \mu^2 p$$

$$= p^2 - \cancel{p^3} + p - 2p^2 + \cancel{p^3}$$

$$= p - p^2$$

$$= p(1-p)$$

$$\boxed{\begin{array}{l} \mu = E(\underline{X}) = p \\ Var(\underline{X}) = p(1-p) \end{array}}$$

b) $\mu = 0$

$Var(\underline{X}) = \sigma^2$

$$Var(2\underline{X}) = E((2\underline{X})^2) - (E(2\underline{X}))^2$$

$$= E(4\underline{X}^2) - (2E(\underline{X}))^2$$

$$= 4E(\underline{X}^2) - 4(E(\underline{X}))^2$$

$$= 4(E(\underline{X}^2) - (E(\underline{X}))^2)$$

$$= 4\,Var(\underline{X})$$

$$= 4\sigma^2$$

$$Var(\underline{X}+3) = E((\underline{X}+3)^2) - (E(\underline{X}+3))^2$$

$$= E(\underline{X}^2 + 6\underline{X} + 9) - (E(\underline{X}+3))^2$$

$$= E(\underline{X}^2) + E(6\underline{X}) + 9 - (E(\underline{X})+3)^2$$

$$= E(\underline{X}^2) + 6E(\underline{X}) + 9 - ((E(\underline{X}))^2 + 6E(\underline{X}) + 9)$$

$$= E(\underline{X}^2) + \cancel{6E(\underline{X})} \cancel{+9} - (E(\underline{X}))^2 \cancel{-6E(\underline{X})} \cancel{-9}$$

$$= E(\underline{X}^2) - (E(\underline{X}))^2$$

$$= Var(\underline{X})$$

$$= \sigma^2$$

$$\boxed{\begin{array}{l} Var(2\underline{X}) = 4\sigma^2 \\ Var(\underline{X}+3) = \sigma^2 \end{array}}$$

7. a) i) $f(n) = \ln(n)$, $g(n) = \log(n)$

$f(n) = O(g(n))$
- let $c=1$ and $n_0 = 1$, then
$0 \leq f(n) \leq g(n)$ for all $n > 1$

$g(n) = O(f(n))$
- let $c = 2$ and $n_0 = 1$, then
$0 \leq g(n) \leq 2 \cdot f(n)$ for all $n > 1$

$\boxed{\text{Both True}}$

ii) $f(n) = 3^n$, $g(n) = n^{10}$

$f(n) = O(g(n))$
- there exists no c, n_0

$g(n) = O(f(n))$
- let $c = 1$ $n_0 \approx 30$, then
$0 \leq f(n) \leq c \cdot g(n)$ for all $n > n_0$

$\boxed{g(n) = O(f(n)) \Rightarrow \text{True}}$

iii) $f(n) = 3^n$, $g(n) = 2^n$
$f(n) = O(g(n))$
- there exists no c, n_0

$g(n) = O(f(n))$
- let $c=1$, $n_0 = 0$, then
$0 \leq g(n) \leq f(n)$ for all $n > 0$

$\boxed{g(n) = O(f(n)) \text{ is True}}$

b) We select the middle element. If it is a 1, we know the transition to be on the LHS. If it is a $\emptyset$, we know the transition to be in the RHS. Then we repeat the process on the appropriate side. At some point, we will get to the transition point. This algorithm works in $O(\log n)$ time because we are essentially halving the work to be done at each iteration, much like a binary search.

8. a)

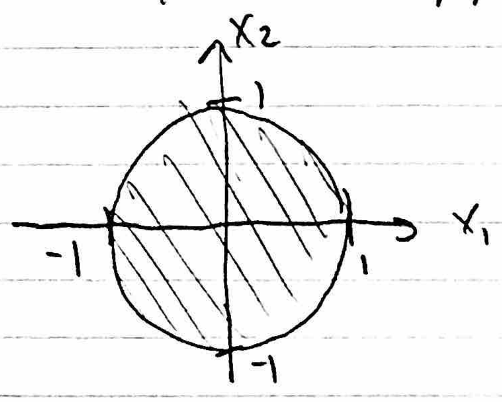
$$
\begin{aligned}
E(\underline{X}\,\underline{Y}) &= \sum_i \sum_j x_i y_j f_{xy}(x_i, y_j) \\
&= \sum_i \sum_j x_i y_j f_x(x_i) f_y(y_j)
\end{aligned}
\quad \Big\} \text{ holds iff } \underline{X} \text{ independent of } \underline{Y}
$$

$$
\begin{aligned}
&= \left(\sum_i x_i f_x(x_i)\right)\left(\sum_j y_j f_y(y_j)\right) \\
&= E(\underline{X}) \cdot E(\underline{Y}) \quad \square
\end{aligned}
$$

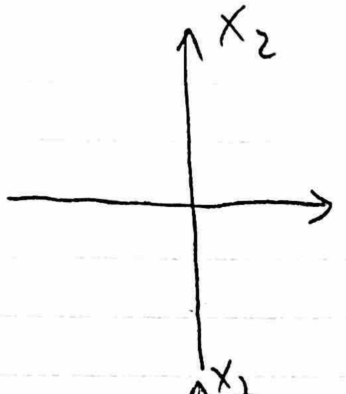
★ b) i) $E(\underline{X}=3) = \frac{1}{6} = \frac{x}{6000}$, $x = 1000$, by LLN

ii) By CLT, the mean of the samples will be equal to the population mean if $n \to \infty$.

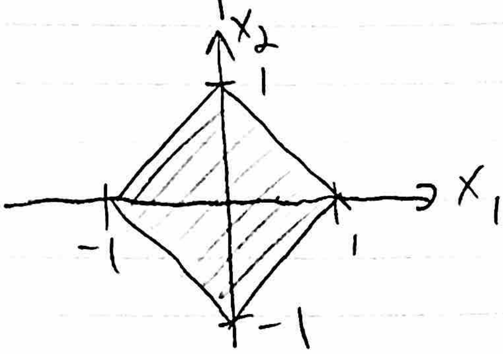
9. i)



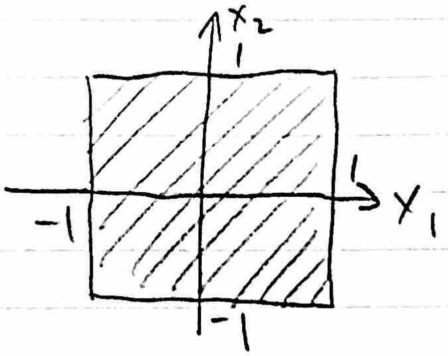
ii)



iii)



iv)



b) i) The eigenvalues are the solutions to the polynomial $\det(A - \lambda I) = 0$.

The eigenvector, v, is a vector corresponding to a particular eigenvalue, λ, such that $(A - \lambda I)v = 0$

ii) $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\det(A - \lambda I) = 0$$
$$\det\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) = 0$$
$$\det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \right) = 0$$
$$(2-\lambda)(2-\lambda) - 1 = 0$$
$$4 - 4\lambda + \lambda^2 - 1 = 0$$

$\lambda^2 - 4\lambda + 3 = 0$

$$(\lambda-1)(\lambda-3)=0$$
$$\lambda=1,3$$

$$(A-\lambda I)v=0$$
$$\left(\begin{bmatrix}2&1\\1&2\end{bmatrix}-\begin{bmatrix}\lambda&0\\0&\lambda\end{bmatrix}\right)v=0$$

let $\lambda=1$

$$\left(\begin{bmatrix}2&1\\1&2\end{bmatrix}-\begin{bmatrix}1&0\\0&1\end{bmatrix}\right)v=0$$

$$\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}v_1\\v_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$=\left[\begin{array}{cc|c}1&1&0\\1&1&0\end{array}\right]$$
$$=\left[\begin{array}{cc|c}1&1&0\\0&0&0\end{array}\right]$$
$$v_1=-v_2$$
$$v=c\begin{bmatrix}-1\\1\end{bmatrix}$$

let $\lambda=3$

$$\left(\begin{bmatrix}2&1\\1&2\end{bmatrix}-\begin{bmatrix}3&0\\0&3\end{bmatrix}\right)v=0$$

$$\begin{bmatrix}-1&1\\1&-1\end{bmatrix}v=0$$

$$=\left[\begin{array}{cc|c}-1&1&0\\1&-1&0\end{array}\right]$$
$$=\left[\begin{array}{cc|c}1&-1&0\\0&0&0\end{array}\right]$$
$$v_1=v_2$$
$$v=c\begin{bmatrix}1\\1\end{bmatrix}$$

$$\boxed{\begin{array}{llll}\text{for} & \lambda=1, & v=c\begin{bmatrix}-1\\1\end{bmatrix} & \text{forall } c\in\mathbb{R}\\ \text{for} & \lambda=3, & v=c\begin{bmatrix}1\\1\end{bmatrix} & \text{for all } c\in\mathbb{R}\end{array}}$$

iii) If λ is an eigenvalue of A, we have

$$Ax=\lambda x$$
$$A\cdot Ax=A\cdot\lambda x$$
$$A^2x=\lambda Ax$$
$$=\lambda\lambda x$$
$$A^2x=\lambda^2x$$

$\Rightarrow$ λ^2 is an eigenvalue of A^2

- we can generalize

$$A^kx=\lambda^kx$$

$\Rightarrow$ λ_i^k is an eigenvalue of A^k, for all $\lambda_i\in$ eigenvalues of A

c) i) $$\frac{d(a^T x)}{dx} = \left[\frac{d}{dx}(a_1 x_1), \frac{d}{dx}(a_2 x_2) \ldots \frac{d}{dx}(a_n x_n) \right]$$
$$= [a_1, a_2 \ldots a_n]$$
$$= a^T$$

$$\boxed{\frac{d}{dx}(a^T x) = a^T}$$

ii) $A = A^T$ by definition of symmetric matrix

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$
$$= \sum_{i=1}^{n} a_{i1} x_i x_1 + \sum_{j=1}^{n} a_{1j} x_1 x_j + \sum_{i=2}^{n} \sum_{i=2}^{n} a_{ij} x_i x_j$$
$$\frac{d}{dx}(x^T A x) = \sum_{i=1}^{n} a_{i1} x_i + \sum_{j=1}^{n} a_{1j} x_j$$
$$= 2 \sum_{i=1}^{n} a_{i1} x_i$$
$$= 2Ax$$
$$\frac{d}{dx}\left(\frac{d}{dx}(x^T A x)\right) = \frac{d}{dx}(2Ax)$$
$$= 2 \frac{d}{dx}(A^T x)$$
$$= 2A^T$$

$$\boxed{\begin{aligned} \frac{d}{dx}(x^T A x) &= 2Ax \\ \frac{d^2}{dx^2}(x^T A x) &= 2A \end{aligned}}$$

d) i) let X_1, X_2 are point $\in$ $w^T x + b = 0$

then, $$\begin{cases} w^T x_1 + b = 0 & (1) \\ w^T x_2 + b = 0 & (2) \end{cases}$$

$$(1) - (2) \Rightarrow w^T (x_1 - x_2) = 0$$

- since $x_1 - x_2$ is a vector on the line $w^T x + b = 0$, and $w^T (x_1 - x_2) = 0$, $X_1 - X_2$ is orthogonal to w^T

- $$w^T (x_1 - x_2) = (x_1 - x_2) w = 0$$ $\square$

ii) $$w^T x + b = 0$$
$$x = \frac{-b}{w^T}$$

distance from origin $= \sqrt{(x_1)^2 + (x_2)^2}$

$$= \sqrt{\left(\frac{-b}{w^T}\right)^2}$$

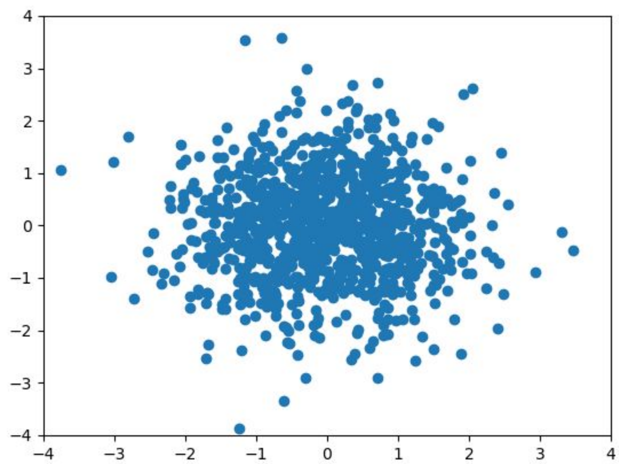
$$= \sqrt{\frac{b^2}{(w^T)^2}}$$

$$= \frac{b}{w^T \cdot w^T} = \frac{b}{\|w^T\|^2} = \frac{b}{\|w\|^2}$$

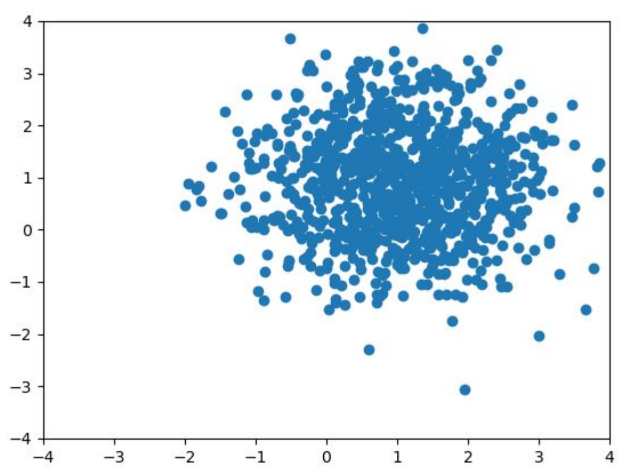
$\square$

10.

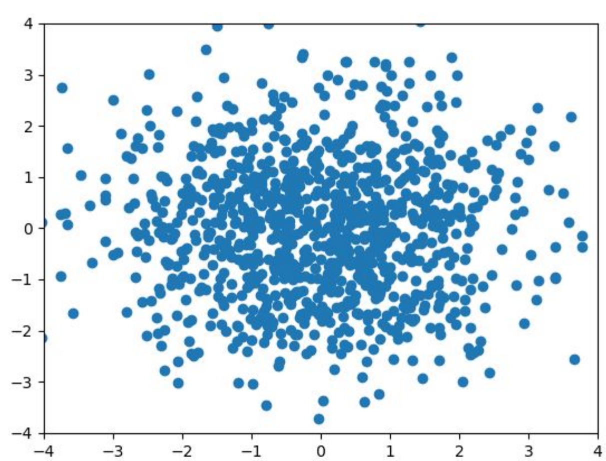
a)


4
3
2
1
0
−1
−2
−3
−4
−4 −3 −2 −1 0 1 2 3 4

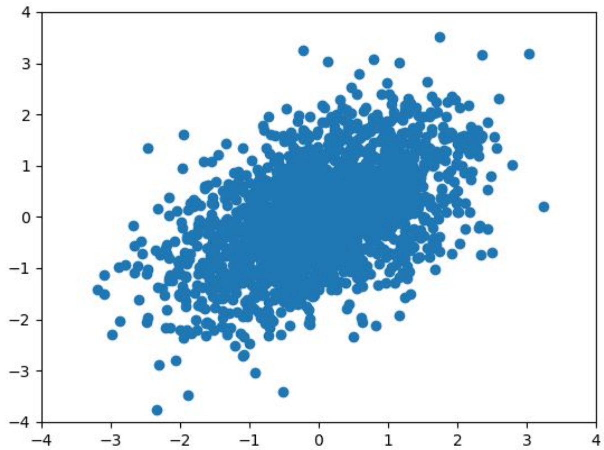
b)


4
3
2
1
0
−1
−2
−3
−4
−4 −3 −2 −1 0 1 2 3 4

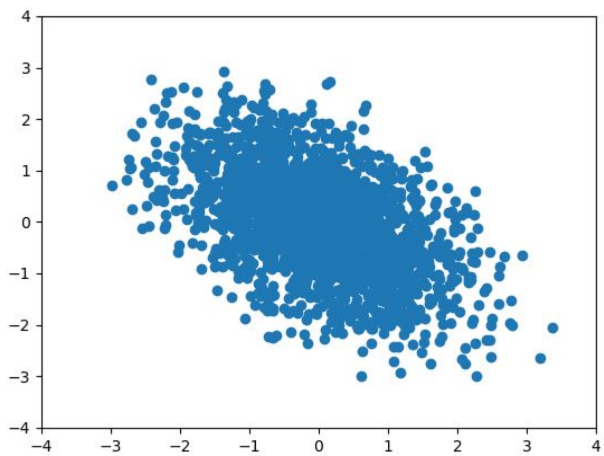
c)


4
3
2
1
0
−1
−2
−3
−4
−4 −3 −2 −1 0 1 2 3 4

d)



e)



11.

eigenvalues: 3, 1
eigenvectors: [0, 0.89442719], [1, -0.4472136]

12.

a) Pima Indians Diabetes Dataset
b) Github
c) This dataset contains data about a particular Pima Indian. The dataset contains medical data of a person and uses those features to predict the likelihood of that person to contract diabetes within the next 5 years.
d) 768

e) 8 - Number of times pregnant, plasma glucose concentration, diastolic blood pressure, triceps skinfold thickness, 2-Hour serum insulin, body mass index, diabetes pedigree function, age